

## PATENT ABSTRACTS OF JAPAN

(11)Publication number : 07-084592

(43)Date of publication of application : 31.03.1995

(51)Int.Cl.

G10L 3/00

G10L 3/00

(21)Application number : 05-228990

(71)Applicant : FUJITSU LTD

(22)Date of filing : 14.09.1993

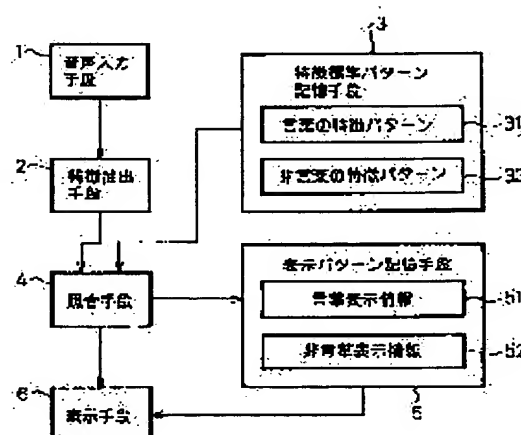
(72)Inventor : IWAMIDA HITOSHI

## (54) SPEECH RECOGNITION DEVICE

## (57)Abstract:

PURPOSE: To recognize a sound other than human words and gives information on it.

CONSTITUTION: The speech recognition device is equipped with a speech input means 1, a feature extracting means 2 which extracts features of a speech signal, a feature standard pattern storage means 3 which stores the feature pattern of the a standard speech signal, a matching means 4 which collates the extracted features of the speech input signal with the stored feature pattern and specifies the standard speech signal corresponding to the input speech signal, a display pattern storage means 5 which stores display information corresponding to the standard speech signal, and a display means 6 which displays the display information; and the feature standard pattern storage means 3 stores the feature patterns 31 of human words and feature patterns 32 other than the human words and the display pattern storage means 5 stores word display information 51 generated by representing the standard speech signal with words as they are and non-word display information 52 corresponding to a non-word standard speech signal.



## LEGAL STATUS

[Date of request for examination] 19.11.1999

[Date of sending the examiner's decision of rejection] 16.07.2002

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's  
decision of rejection]

[Date of extinction of right]

(19)日本国特許庁 (J P)

(12) 公 開 特 許 公 報 (A)

(11)特許出願公開番号

特開平7-84592

(43)公開日 平成7年(1995)3月31日

(51)Int.Cl.<sup>6</sup>

G 1 0 L 3/00

識別記号

5 3 1 N

庁内整理番号

9379-5H

5 6 1 H

9379-5H

F I

技術表示箇所

審査請求 未請求 請求項の数 6 O L (全 7 頁)

(21)出願番号 特願平5-228990

(22)出願日 平成5年(1993)9月14日

(71)出願人 000005223

富士通株式会社

神奈川県川崎市中原区上小田中1015番地

(72)発明者 岩見田 均

神奈川県川崎市中原区上小田中1015番地

富士通株式会社内

(74)代理人 弁理士 宇井 正一 (外4名)

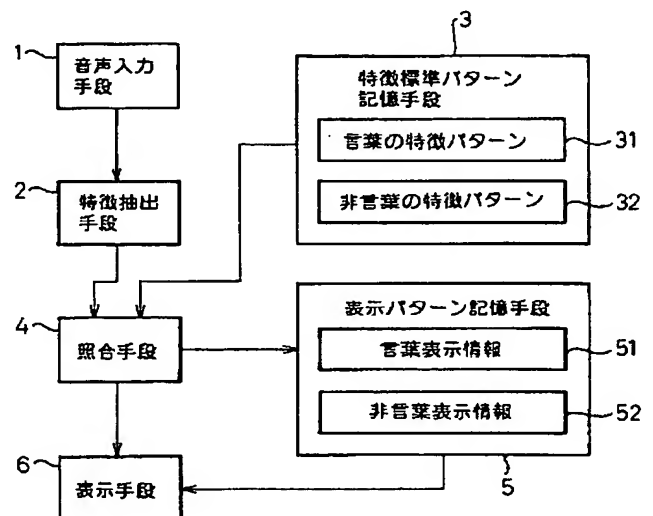
(54)【発明の名称】 音声認識装置

(57)【要約】

【目的】 本発明は音声認識装置に関し、人間の言葉以外の音を認識して報知できるようにすることを目的とする。

【構成】 音声入力手段1と、音声信号の特徴を抽出する特徴抽出手段2と、標準音声信号の特徴パターンを記憶する特徴標準パターン記憶手段3と、抽出された音声入力信号の特徴と記憶された特徴パターンとを照合し入力音声信号に対応する標準音声信号を特定する照合手段4と、標準音声信号に対応する表示情報を記憶する表示パターン記憶手段5と、表示情報を表示する表示手段6とを備える音声認識装置において、特徴標準パターン記憶手段3は、人間の言葉の特徴パターン31と人間の言葉以外の特徴パターン32を記憶しており、表示パターン記憶手段5は、標準音声信号をそのまま文字にした言葉表示情報51と、非言葉標準音声信号に対応する非言葉表示情報52とを記憶しているように構成する。

本発明の原理構成図



1

## 【特許請求の範囲】

【請求項1】 音声信号を入力する音声入力手段（1）と、

音声信号を認識するための特徴を抽出する特徴抽出手段（2）と、

標準音声信号の特徴パターンを記憶する特徴標準パターン記憶手段（3）と、

抽出された音声入力信号の特徴と前記特徴標準パターン記憶手段（3）に記憶された特徴パターンとを照合し、入力された音声信号に対応する標準音声信号を特定する照合手段（4）と、

前記標準音声信号に対応する表示情報を記憶する表示パターン記憶手段（5）と、

該照合手段（4）で標準音声信号が特定された場合に、標準音声信号に対応する表示情報を表示する表示手段（6）とを備える音声認識装置において、

前記特徴標準パターン記憶手段（3）は、人間の言葉の標準音声信号の特徴パターン（31）の他に、人間の言葉以外の非言葉標準音声信号の特徴パターン（32）を記憶しており、

前記表示パターン記憶手段（5）は、前記標準音声信号をそのまま文字にした言葉表示情報（51）と、前記非言葉標準音声信号に対応する非言葉表示情報（52）とを記憶していることを特徴とする音声認識装置。

【請求項2】 前記非言葉表示情報（52）は画像情報であり、

前記表示手段（6）は、画像情報を表示可能であることを特徴とする請求項1に記載の音声認識装置。

【請求項3】 前記非言葉表示情報（52）の画像情報は動画画像情報を含み、前記表示手段（6）は動画画像情報を表示可能であることを特徴とする請求項2に記載の音声認識装置。

【請求項4】 前記表示手段（6）に情報が表示されたことを、音声以外の手段で報知する報知手段を備えることを特徴とする請求項1に記載の音声認識装置。

【請求項5】 前記照合手段（4）は、抽出された音声入力信号の特徴と前記特徴標準パターン記憶手段（3）に記憶された特徴パターンとの一致具合を順位付けて出力するものであり、

前記表示手段（6）は、所定の一致具合の標準音声信号までの複数の表示情報を表示することを特徴とする請求項1に記載の音声認識装置。

【請求項6】 前記照合手段（4）は、同時に複数の標準音声信号が入力されたことを検出した場合には、前記表示手段（6）は複数の表示情報を表示することを特徴とする請求項1に記載の音声認識装置。

## 【発明の詳細な説明】

## 【0001】

【産業上の利用分野】本発明は、入力された音声信号を自動認識し、その認識結果を表示する音声認識装置に関

2

し、特に人間の言葉以外の音声信号も認識しその音声信号が認識できる形で表示するようにした音声認識装置に関する。

## 【0002】

【従来の技術】音声認識装置が実用化されつつある。従来の音声認識装置は、あらかじめ認識すべき音声信号を標準パターンとして登録して記憶し、入力された音声信号とこれらの標準パターンとを照合し、もっとも良く一致する標準パターンを特定し、その一致具合が所定のレベル以上の場合に標準パターンの音声が入力されたとして認識している。標準パターンとして登録するためには、その標準パターンを入力する必要があるが、通常標準パターンとしては使用者が入力できる音声信号に限られるが、ブラント等の自動制御のために異常音を認識する場合には、そのような異常音を標準パターンとする場合もある。

【0003】認識結果を、音声信号の認識が正しく行われたものとしてそのまま出力する場合もあるが、認識の正確さを期すために、認識結果に従って合成した音声を出力し、話者に確認してもらうことが一般に行われる。この確認方法には、合成音声を出力するだけでなく、認識結果を文字情報として表示して確認する方法もあるが、いずれにしろ、入力された音声信号を正確に表現できる方法であることが必要である。そのため、人間の話言葉を標準パターンとし、確認のための音声合成又は文字表示が行えるようにしている。上記の異常音を標準パターンとした場合には、確認する必要がないため確認は行われない。

【0004】聴覚障害者が他者と会話する場合には手話や筆談によるが、上記の音声認識装置を利用して聴覚障害者が通常の言語能力を有する者と会話するのを支援することが考えられる。通常の言語能力を有する者は、通常通り会話を行い、その音声を音声認識装置で認識し、認識結果を文字情報として表示して聴覚障害者に知らせる。

【0005】本発明は、特にこのような目的で使用される音声認識装置に適しているが、このようなものに限られず、例えば、会議中のような音を発することが好ましくない状態で、別の場所の音声情報を知る必要がある場合等に使用される音声認識装置にも適用可能である。しかしここでは、聴覚障害者用音声認識装置が例としてもっとも適当と思われるので、これを例として説明する。

## 【0006】

【発明が解決しようとする課題】上記のような認識した音声信号を表示する音声認識装置では、認識した音声信号が表示できることが必要であるから、登録される標準パターンは人間の話言葉に限定されていた。また、表示するのは人間の話言葉に対応した文字であるから、表示装置も文字を表示できるものであればよかった。

【0007】聴覚障害者用音声認識装置が通常の言語能

力を有する者との会話を支援するためだけのものであれば、上記のような登録される標準パターンは人間の話言葉に限定され、表示装置も文字を表示できる音声認識装置で十分と思われる。しかし、聴覚障害者により広い音声情報を提供するという観点からは、人間の話言葉以外の音声情報も提供できることが望ましい。

【0008】本発明は、このような観点から、人間の話言葉以外の音声情報も表示情報として提供できる音声認識装置の実現を目的とする。

【0009】

【課題を解決するための手段】図1は本発明の音声認識装置の原理構成図である。図示のように、本発明の音声認識装置は、音声信号を入力する音声入力手段1と、音声信号を認識するための特徴を抽出する特徴抽出手段2と、標準音声信号の特徴パターンを記憶する特徴標準パターン記憶手段3と、抽出された音声入力信号の特徴と特徴標準パターン記憶手段3に記憶された特徴パターンとを照合して入力された音声信号に対応する標準音声信号を特定する照合手段4と、標準音声信号に対応する表示情報を記憶する表示パターン記憶手段5と、照合手段4で標準音声信号が特定された場合に標準音声信号に対応する表示情報を表示する表示手段6とを備える音声認識装置において、上記目的を達成するため、特徴標準パターン記憶手段3は、人間の言葉の標準音声信号の特徴パターン31の他に、人間の言葉以外の非言葉標準音声信号の特徴パターン32を記憶しており、表示パターン記憶手段5は、標準音声信号をそのまま文字にした言葉表示情報51と、非言葉標準音声信号に対応する非言葉表示情報52とを記憶していることを特徴とする。

【0010】

【作用】本発明の音声認識装置において、特徴標準パターン記憶手段3に、人間の言葉以外の音声信号、例えば、「サイレン」の音を非言葉標準音声信号の特徴パターン32として記憶しておけば、サイレンの音を認識することができる。そして、サイレンの音を認識した時には、表示パターン記憶手段5にサイレンの音に対応する非言葉表示情報52として記憶されている「サイレンが鳴っています。」という表示を行うことができる。

【0011】また、「サイレンが鳴っています。」というだけでは、他者が「サイレンが鳴っています。」と言っているのか、実際に「サイレン」の音が聞こえるのか判別できないため、例えば、消防車の「サイレン」の音であれば、消防車の画像を表示するようにすれば、より大きな情報を提供することが可能になる。

【0012】

【実施例】図2は本発明の1実施例の音声認識装置の構成を示す図であり、聴覚障害者のための音声認識装置である。図2において、参照番号11は音声信号を電気信号に変換するマイクロフォン、21はマイクロフォン11から得られる電気信号を、例えば、12kHzのサン

プリング周期でサンプリングしてデジタル信号に変換するA/D変換器、22は周波数分析部である。周波数分析部22は、A/D変換されたデジタル時系列信号を10ms毎にファーストフーリエ変換(FFT)等で周波数分析し、聴覚的な尺度で8帯域に分割した各周波数帯域でのパワーを求め、周波数特徴パラメータの時系列を得る。33は標準パターン格納用ROMであり、上記のマイクロフォン11、A/D変換器21及び周波数分析部22であらかじめ求めておいた認識対象信号の周波数特徴パラメータの時系列を記憶する。ここでは、認識対象信号の周波数特徴パラメータの時系列を標準パターンと呼ぶことにする。41は照合部であり、DPマッチング等の手法を用いて、入力音声信号の周波数特徴パラメータの時系列(入力パターン)と標準パターンとの照合を行い、入力パターンともっともよく一致する標準パターンを求める。53は文字の表示パターンを記憶したキャラクタ用ROMであり、53は絵等のグラフィック表示パターンを記憶したグラフィックパターン用ROMであり、61は表示制御部であり、62はCRT、液晶表示装置等の表示装置である。

【0013】標準パターン格納用ROM33に記憶される標準パターンの種類は、標準パターン格納用ROM33の容量及び照合部41の処理能力によって決定される。標準パターン格納用ROM33の容量を増加させれば、記憶できる標準パターンの種類は増加するが、入力パターンとの照合に要する演算量が増加するため、標準パターンの種類が増加した場合には、音声信号が入力されてから認識して表示するまでの時間が長くなる。そのため、標準パターンの種類は照合部41の処理能力を考えて決定される。本実施例では、標準パターンとして、日常生活に必要な「おはよう」、「食事です」等の数十種類の言葉と、消防車のサイレン音、乳児の泣き声等の数種類の言葉以外の音が記憶される。

【0014】照合部41は実際にはコンピュータであり、DPマッチングの手法を用いて入力パターンに近い標準パターンを探し出す。ここで、DPマッチングの手法について簡単に説明する。図3はDPマッチングの概念を示す図である。図中、 $a_1$ 、 $a_2$ 、 $a_3$ …は入力の特徴パラメータ時系列、 $b_1$ 、 $b_2$ 、 $b_3$ …は標準パターンの周波数特徴パラメータ時系列である。DPマッチングでは、入力と標準パターン周波数特徴パラメータ時系列がもっともよく一致するように時間軸を変換した後、距離を求める。すなわち、図中の経路Pが最適な経路とすると、そのP上の各位置で対応するaとbの差をP上の全位置について総和したものを入力と標準パターンの距離とする。このようにして全標準パターンについての距離を求め、もっとも距離の小さい標準パターンを認識結果とする。

【0015】グラフィックパターン用ROM54には、各標準パターンのうち、非音声の標準パターンに対応す

る画像情報、すなわち、消防車の絵や乳児が泣いている絵等が標準パターンに対応付けて記憶されている。表示制御部 61 は、照合部 41 での照合結果に従って、もっともよく一致する標準パターンに対応する表示を表示装置 62 に行う。表示する標準パターンが音声であった場合には、表示する文字列を表すコードと各文字のパターンをキャラクタ用 ROM 53 から読み出して合成し、表示装置 62 にビデオ信号として出力する。すなわち、入力された音声信号のこばを表す文字が表示される。なお、キャラクタ用 ROM 53 に標準パターンに対応する文字列のパターンを記憶させておくことも可能であり、そのようにすれば、表示制御装置 61 を簡単にすることができる。

【0016】入力された音声信号がこばでない時には、その音が聞こえることを示す文字情報か、グラフィック表示パターンを表示する。図 4 はその例である。図 4 の (1) は、入力された音声信号がこばでない時に、その音が聞こえることを文字で表示する例であり、(2) は関連させたグラフィック表示パターンを表示する例である。どちらも、(a) がサイレンの音が聞こえる場合を示し、(b) が赤ん坊が泣いている場合を示す。前述のように、文字で表示した場合には、他者がそのような言葉を言っている場合との差が判別できないという問題があるので、入力された音声信号がこばでない時にはグラフィックパターンが表示されると決めておくか、図 4 の (1) に更に、他者がそのような言葉を言っていないことを示す表示を追加することが望ましい。

【0017】以上が本発明の実施例であるが、照合部 41 における照合は、図 3 に示すように、入力した周波数特徴パラメータ時系列と標準パターンの周波数特徴パラメータ時系列の経路 P に対して各位置で対応する a と b の差を P 上の全位置について総和したものを距離とした。しかし、全位置について総和したものを距離とするのではなく、図 5 に示すように、経路 P の始端と終端の入力パターンに対応する位置をそれぞれ  $a_1$  と  $a_1$  に固定せず求めるといった変形例も可能である。この図の例の場合は、 $a_m$  と  $a_n$  がそれぞれ始端と終端になっている。このようにして、入力パターンの一部分との距離がもっとも小さい標準パターンを求めてそれを認識結果とする。

【0018】音声認識する対象の音をマイクロフォンで捕らえるが、複数の音源が存在するために、マイクロフォンが複数の音を同時に捕らえる場合がある。このような場合、そのまま標準パターンと照合したのでは一致する標準パターンを特定するのは難しくなる。このような場合には、例えば、指向性のマイクロフォンを複数、指向性を異ならせて設け、各マイクロフォンの入力信号を比較して、一致する場合には音源は 1 個であると認識して照合処理を行い、各マイクロフォンの入力信号が異なる場合には異なる音源が存在すると認識し、それぞれの

音源からの音声信号について照合処理を行う。

【0019】複数の音声信号を認識した場合には、同時に認識した音声信号を表示する必要がある。そこで、例えば、人間の言葉と消防車のサイレンを認識した場合には、人間の言葉を文字表示すると共に、消防車の画像を表示する。また、消防車のサイレンと赤ん坊の泣き声を認識した場合には、消防車の画像と赤ん坊が泣く画像を同時に表示する。図 6 の (1) はこの表示例を示している。

【0020】更に、照合部 41 における照合の結果、距離の類似した標準パターンが複数存在する場合があり得る。このような場合、その中のもっとも近い標準パターンを特定するのは危険である。そこでこのような場合には、可能性のある複数の標準パターンをそのまま表示するといった具合に、表示を変形する。図 6 の (2) は、そのような変形例における表示例を示す図であり、赤ん坊の泣き声か猫の鳴き声か完全に特定できない場合に、赤ん坊が泣いている画像と猫の画像を同時に表示し、認識結果に応じてそれぞれに確率を表示している。

【0021】以上のように、上記の装置を使用すれば人間の話言葉以外の音も認識して関連する情報を表示することができるようになる。話言葉を認識して表示するのであれば、相手に対面して会話する時だけに使用するものであるから、その時に表示画面を見れば良いが、人間の話言葉以外の音はいつ入力されるかわからない。そのため、装置は常時オン状態とし、そのような音が入力されたことを認識した時に表示を行うことになるが、装置を使用している人はそのような音の入力があっても認識できないと考えられる。しかも常時表示画面を見ていることはできない。そこで、人間の話言葉以外の特定の音が入力されことを認識した時には、音や表示以外の別の手段、例えば、振動等により、使用者に報知するようにする。このような装置を聴覚障害者の支援装置に使用し、消防車のサイレンを認識した時に振動により報知するようにすれば、聴覚障害者の安全に大いに寄与する。

#### 【0022】

【発明の効果】以上説明したように、本発明によれば、音声認識装置において、人間の言葉だけを認識して文字で表示するだけでなく、人間の言葉の音も認識してそれに対応する情報を表示できるようになる。このような装置を聴覚障害者用に使用した場合、聴覚障害者に周囲の音も含めた情報を提供できるようになる。

#### 【図面の簡単な説明】

- 【図 1】本発明の音声認識装置の原理構成図である。
- 【図 2】本発明の実施例の装置の構成を示す図である。
- 【図 3】DP マッチングの説明図である。
- 【図 4】非言葉の表示パターンの例を示す図である。
- 【図 5】DP マッチングの変形例の説明図である。
- 【図 6】非言葉の表示パターンの変形例を示す図である。

## 【符号の説明】

1…音声入力手段

2…特徴抽出手段

3…特徴標準パターン記憶手段

4…照合手段

5…表示パターン記憶手段

6…表示手段

3 1…言葉の特徴パターン

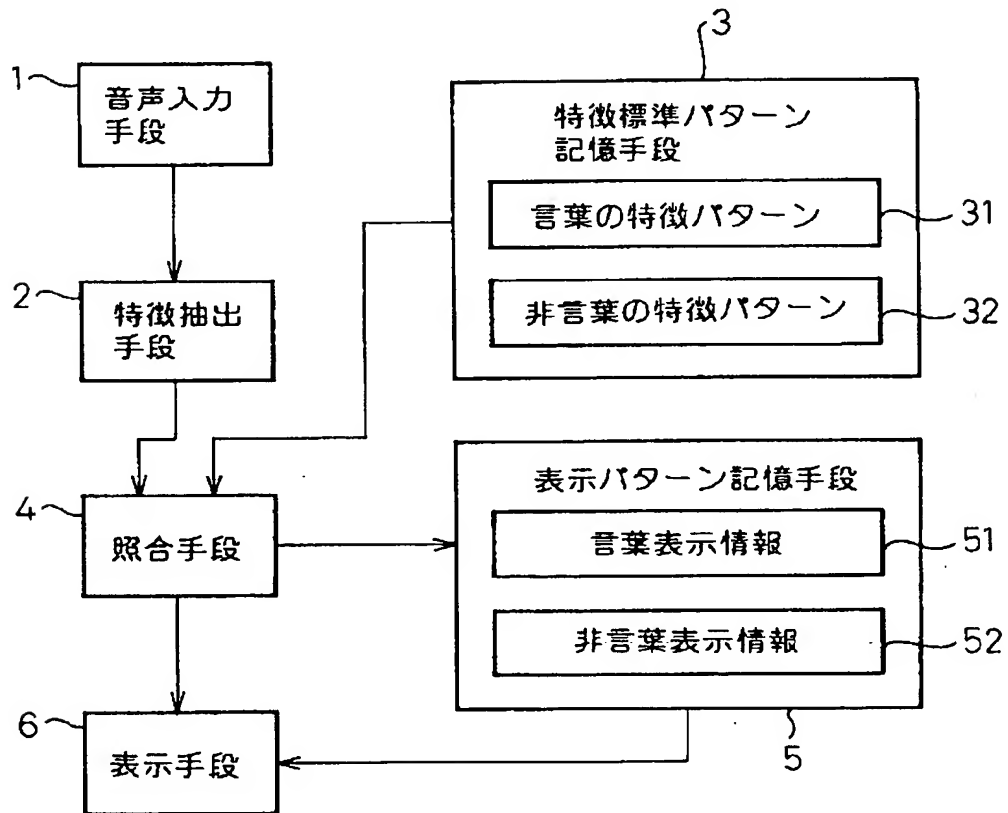
3 2…非言葉の特徴パターン

5 1…言葉表示情報

5 2…非言葉表示情報

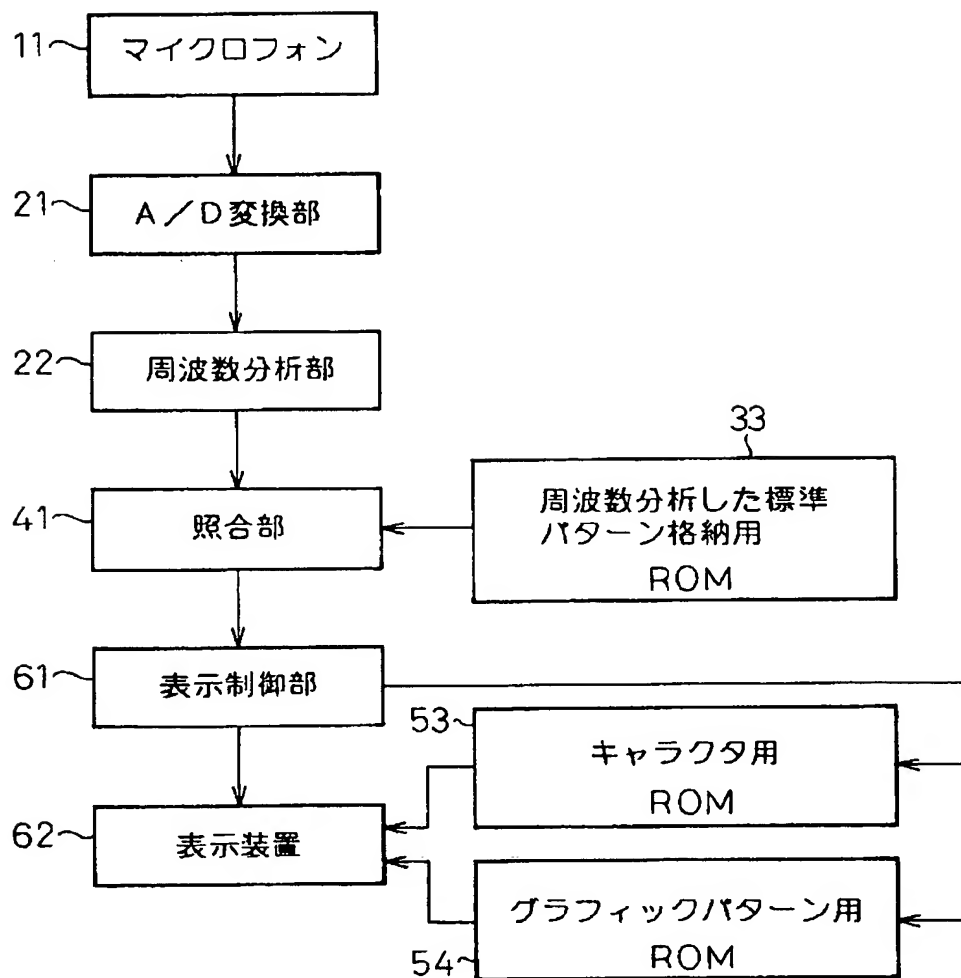
【図 1】

## 本発明の原理構成図



【図2】

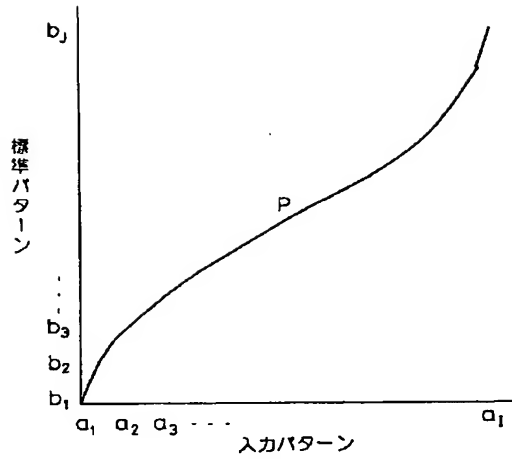
## 実施例の装置の構成





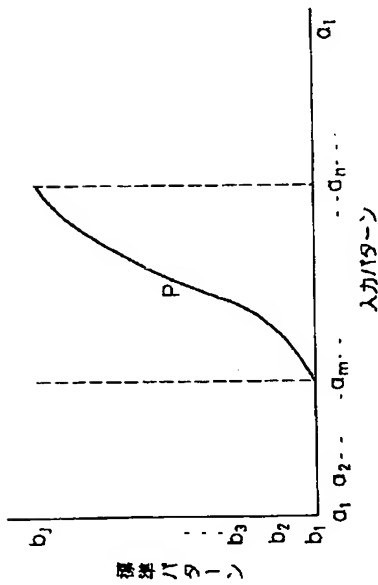
【図 3】

DPマッチングの説明図



【図 5】

DPマッチングの変形例



【図 4】

非言葉の表示パターン

## (1) 文字表示

## (a) サイレンの音

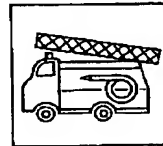
サイレンが鳴っています。

## (b) 赤坊の泣き声

赤ちゃんが泣いています。

## (2) グラフィック表示

## (a) サイレンの音



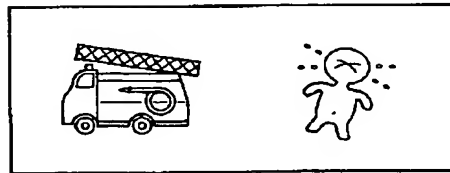
## (b) 赤坊の泣き声



【図 6】

表示の他の例

## (1) 同時に2種の音



## (2) 類似の音を完全に識別できない時

